PREDICTIVE MODELLING OF CYSTIC PANCREATIC NEOPLASM HISTOLOGY

By

Naveed A. Pasha MD

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master Science

Baltimore, Maryland

March, 2017

Project Abstract

Introduction: Cystic pancreatic tumours constitute the second-most prevalent group of tumours after ductal adenocarcinoma of the pancreas. They are broadly categorized histopathologically as mucinous and non-mucinous. Mucinous cystic tumours, which comprise intraductal papillary mucinous neoplasm (IPMN) and mucinous cystic neoplasm (MCN), display a greater predisposition for conversion to malignancy as compared to their non-mucinous counterpart serous cystadenoma (SCN). Typically, cystic tumours are diagnosed when they became symptomatic, owing to growth in size or malignant transformation whereby invasion of gastric wall and splenic vein leads to bleeding gastric varices.

Problem: With the advent of modern imaging, cystic tumours are being detected incidentally with increasing frequency at a stage when they are very small, asymptomatic and unlikely to have malignant transformation. Indiscriminate pre-emptive resection of cystic tumours with malignant potential is not feasible owing to the great morbidity associated with resection procedures. Hence patients with cystic tumours of malignant potential need to be carefully chosen for resection. There exists no single diagnostic tool that can accurately discriminate between mucinous and non-mucinous cystic pancreatic neoplasms except for histopathology which is only possible after resection of tumour. Consensus guidelines exist for non-invasive identification of tumour histological category using radiological features as surrogates. The process of diagnosis remains dependant on expert opinion. The goal of this study is to harness demographic and radiological features with potential capacity to discriminate among cystic

pancreatic neoplasms, as reported in consensus guidelines and literature, to form a simplified predictive tool.

Methods: Excluding children and adolescents, all cases of cystic masses, with a histopathological diagnosis confirmed as SCN, IPMN or MCN, between 1995 to date were searched in a prospectively maintained database. Data on symptoms at presentation (if any) and computed tomography features (CT), magnetic resonance imaging features (MRI), endoscopic ultrasound features (EUS) were stored. A statistical tree based algorithm was run on the data to develop a decision tree that will predict for the three types of cystic neoplasms. The performance of the decision tree was assessed using a receiver operating curve surface (ROCS) and volume under the surface (VUS).

Results: The VUS for the decision tree was 77.5% which means that the decision tree correctly classified SCNs, IPMNs and MCNs in 77.5 % of all cases. The empirical 95% confidence interval for the VUS was 54.5% - 78.5%. of 36.2% of SCNs that were misclassified as IPMNs or MCNs and 15.9% of non-SCN cases that were misclassified as SCNs.

Conclusion: The decision tree offers a fairly accurate, yet simplified, means of predicting for the histology of cystic pancreatic lesions and is therefore capable of supporting decisions in clinical management of these patients. The accuracy of the decision tree may be raised higher by incorporation of biochemical marker and molecular marker data from cyst fluid aspiration.

Preface

The author acknowledges the following for their contribution in collection of data necessary for this paper, analysing the data to achieve appropriate results and interpreting the results in context of the clinical problem being addressed:

- Zheyu Wang PhD[2], [2]Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Centre, Johns Hopkins University School of Medicine, Baltimore, MD, USA
- Ammar A. Javed MD[3], Department of Surgery, The Johns Hopkins Hospital, Baltimore, MD, USA
- Harold Lehmann MD PhD, Division of Health Sciences Informatics, Johns Hopkins University School of Medicine, Baltimore, MD, USA
- Jin He MD PhD, Department of Surgery, The Johns Hopkins Hospital, Baltimore, MD, USA

Table of Contents

List of Tables

Table of Figures

1. Introduction

1.1 General Information

The WHO classifies pancreatic tumours histopathologically according to cell of origin followed by malignant potential. The most frequent pancreatic tumours are those that arise from the epithelial cells of the pancreas of which the most common is ductal adenocarcinoma, a solid malignant tumour that is traditionally known to be responsible for 85% of all pancreatic tumours. (1) The remainder of the epithelial tumours are cystic neoplasms of varying malignant potential. True incidence of these cystic neoplasms is hard to estimate considering the rarity of the disease as well as the fact that not all cystic tumours may become symptomatic such that necessitates presentation to a healthcare facility. Early surgical series that first attempted to characterise these tumours are characterised by a small cohort of patients (2) and Kimera et al., in an autopsy series of 300 cases, identified the presence of cystic neoplasms in 73 cases (24.3%) where the tumour was not the probable cause of death. (3)

The most notable tumours in the cystic category in terms of incidence are intraductal papillary mucinous neoplasms (IPMN), serous cystadenoma (SCN) and mucinous cystic neoplasms (MCN) in that order.

Serous cystic adenomas are benign tumours that appear histopathologically as “glycogen-rich, ductular-type epithelial cells that produce a watery fluid similar to serum”. (1)These cells are found lining fibrous septae that are rich in capillary vessels. Often the fibrous septae will converge to a dense fibrous core which is seen grossly as a stellate scar. (4) Normally, such

septations give the tumour the appearance of possessing many small cysts whereupon the tumour is labelled as microcystic. However, there is another manifestation of SCNs which contains larger (1 – 2 cm in diameter) and fewer cysts that may protrude into the pancreatic parenchyma, labelled as the oligocystic variant also known as the macrocystic variant. Seventy percent of cases are in females when both variants are taken into account (1) and this propensity is largely driven by the affinity of the more common microcystic variant to that gender; the oligocystic variant is seen equally in both genders. (5) SCNs frequently occur in the body or tail of the pancreas (1) although, once again, this is largely because the more common microcystic variant is invariable found in these regions. The macrocystic variant is more commonly observed in the head of the pancreas where it may exert mass effect on the biliary tree resulting in obstructive jaundice. (6)

Mucinous tumours, as the name asserts, are characterised by excessive production of mucus which distinguishes them from SCNs. Intraductal papillary mucinous neoplasms (IPMN) are defined by Klöppel et al. as category of "tumours that are grossly visible, mucin-producing, cystic tumours growing within the pancreatic ducts and forming papillary projections". They occur within the pancreatic ducts and are therefore classified according to whether they occur in the main pancreatic duct or branch pancreatic ducts. Columnar mucin-containing cells line the ducts and form pseudopapillary structures. (1) The excessive production of mucus combined with the mass effect of the pseudopapillary structures usually leads to dilation of the duct distal to the location of the IPMN. Therefore they are not truly cystic but appear cystic due to the pseudopapillary structures protruding into the dilated pancreatic duct, interspersed by mucus. When the IPMN occurs in the main duct, the portion of the duct distal to the tumour

is seen to be markedly dilated, often greater than 1 cm, on gross appearance. However, when the IPMN occurs in the branch duct this gross dilation does not occur and the tumour instead takes the form of a singular cysts or a cluster of cysts, often referred to as a "cluster of grapes" appearance. (7) IPMNs are seen slightly more in more in males with a female to male ratio of 1:1.5 and are 80% of instances are observed in the head of the pancreas. These tumours gradually undergo increasing dysplastic changes to eventually become invasive carcinoma of the pancreas. (8)

Similarly, mucinous cystic neoplasms (MCN) are also composed of columnar epithelium that produces mucin, progresses along degrees of dysplastic changes to eventually become invasive carcinoma (5) and additionally is supported by an ovarian-like stroma. (1) MCNs are typically multilocular and fairly large at the time of diagnosis (60 – 100 mm). Furthermore, the wall of the cysts has mural nodules and papillary projections towards the lumen except in case of the more infrequent unilocular variants which have smooth walls. (8) In contrast to IPMNs, however, they do not communicate with the pancreatic ductal system with a proclivity towards the body and tail regions of the pancreas. Additionally, MCNs are found almost exclusively in women over a broad age range excepting childhood and adolescence. (5)

Radiological imaging reflects the differences in gross appearance between the kinds of pancreatic cysts. CT shows serous cystadenomas as a collection of six or more small cysts, again up to 2 cm in size, with a "spongy" appearance displaying soft-tissue or mixed attenuation with enhancement of septae. SCNs may or may not display the classical central stellate scar. The tumours are often well marginated and the boundaries may be smooth or lobulated. The oligocystic variant may be difficult to differentiate from a mucinous cystadenoma as it may

appear as a large unilocular cysts or may contain fewer large cysts > 2cm. MCNs appear as multilocular, macrocystic lesions (> 2cm) lesions with smooth boundaries that may have peripheral calcification and that may contain septations. Their large size may cause partial obstruction of the pancreatic ductal system and so dilated ducts may be seen. (9,10) Main duct IPMNs can mimic an MCN if it involves only a segment of the main pancreatic duct or may manifest as diffuse dilation when the entire main pancreatic duct is involved. In either case, papillary nodules may be seen protruding from the walls and the walls appear thickened and enhance on contrast. Branch-duct IPMNs appear as a lobulated collection of cysts of any number around a local branch duct dilation and normal main pancreatic duct calibre. (11)

Endoscopic ultrasound (EUS) findings are interpreted in conjunction with CT images on the basis of morphology and cyst fluid analysis. (12) The main features of interest are the size of cysts, the shape of the margins of cysts and the walls of the cysts. Serous cystadenomas continue the trend of appearing as lobulated, multiloculated lesions with an occasional central scar. They may display echogenic foci or may appear as a hypoechoic cystic/solid mass. (13) The exception is that of the macrocystic variant which appears remarkably similar to MCN on CT and MRI in that both are unilocular. However, on EUS, MCNs display thick walls and peripheral calcifications and loculations whereas the unilocular SCNs display microcystic components of less than 2 mm in size. (14,15) Pancreatic ductal dilation idiosyncratic to IPMNs is also seen in EUS along with thickened hyperechoic walls of the dilated ducts accompanied with hypoechoic mass within with solid and cystic components of variable size. (13)

EUS affords the opportunity to aspirate fluid within the cysts which also reveals differences when analysed for brush cytology, pancreatic enzymes and tumour markers. (7) Aspirates from

serous cystadenomas are typically acellular whereas mucinous cysts have a good cellular yield. When yield is obtained, cells can be assessed on brush cytology according to cellular descriptions mentioned previously. Carcinoembryonic antigen (CEA) levels are generally higher in mucinous lesions than serous lesions. Brugge et al. report a mean and median of 5607 ng/mL and 500 ng/mL for mucinous lesions as opposed to 284 ng/mL and 21 ng/mL for serous lesions respectively. This difference in CEA concentration is corroborated by a multicentre review by Van der Waaij et al. who observed that a CEA > 800 ng/mL is strongly suggestive of mucinous cystadenoma and a CEA < 5 ng/mL is strongly suggestive of serous cystadenoma as is a Cancer Antigen 19-9 < 37 ng/mL. Similarly, both studies found that cyst fluid from mucinous lesions have higher amylase and lipase levels than serous lesions. (16,17)

MRI shows serous cystadenomas to be a cluster of small cysts ("honeycomb pattern") each of which range between 0.1 – 2 cm in size. The exception is the oligocystic variant whose cysts are fewer in number and appear larger. These cysts do not communicate with the pancreatic ductal system, their walls enhance in the delayed phase of contrast enhanced T1-weighted images and their lumens display simple fluid signal intensity on T2-weighted images which is in keeping with the presence of serous fluid. As the lesion grows as a whole, fibrous tissue retraction distorts the architecture of central cysts which may collapse to form a central scar. This scar may become calcified leading to a signal void on MRI. (18) The varying amounts fibrous tissue seen across cysts means that ADC values on diffusion-weighted images will vary, making the latter a poor choice for diagnosis. (19)

MCNs usually appear, on MRI, as round or lobular uniloculated lesions with low signal intensity on T1-weighted imaging and high signal intensity with T2-weighted imaging. Cysts are often

seen to be larger than 2 cm and a few septae may be seen when the lesion is multiloculated (20%). The walls show enhancement on delayed gadolinium enhanced, fat-supressed T1-weighted images, reflecting fibrotic changes, as well as high ADC values and 16% of cases may exhibit peripheral calcifications. IPMNs have a varied appearance on MRI depending on the type that is found. In all cases, a clear communication with the pancreatic ductal system is clearly visualised in the background of pancreatic parenchyma that has generally low signal intensity reflective of atrophic changes. If the main pancreatic duct is involved then, as in CT, the dilation may be diffuse or segmental according to extent of involvement. If branch duct is involved, then only local ductal dilation is observed on T2-weighted images. ADC values are elevated on diffusion-weighted imaging along the length of the duct involved due to the presence of mucous. The tumour itself is usually no visualised; its presence is inferred from the multifocality of the lesions observed. (18–20)

1.2 Rationale & Background Information

Ganjoux et al., in their retrospective review of a prospectively maintained registry of 1424 patients diagnosed with pancreatic cysts, show that cystic pancreatic neoplasms are being detected with greater frequency with the advent of non-invasive imaging techniques. Furthermore, they are being detected at a stage where they are small in size and when they generally lack features indicative of malignancy. (21) Incidentally-detected cystic pancreatic neoplasms now pose a significant diagnostic and management challenge owing to the lack of information of their natural history and diagnosis at very early stage. (22) While the WHO classification helps surgeons understand the malignant potential of each tumour, there is no information as to why and at what point in time these tumours transform to malignancy. It

appears that SCNs, MCNs and IPMNs progress at different rates, along an adenoma carcinoma sequence and the malignant end of the spectrum manifests as cystadenocarcinoma for the former two and invasive carcinoma for the latter. The prognosis of these malignant forms depends on the degree of spread. (1)

A variety of surgical procedures may be employed for resection of cystic neoplasms depending on location of the tumour and secondarily on the size of the tumour. These include Whipple procedure, total pancreatectomy, segmental resection, enucleation and distal pancreatectomy that may or may not be spleen preserving. (21,23) Generally these procedures carry great morbidity in the post-operative period (43%) exhibiting complications such as cardiopulmonary event, fistula, delayed gastric emptying and sepsis.(24)

The lack of aggressiveness typically seen in cystic neoplasms of the pancreas coupled with their high rate of serious morbidities seen in surgical procedures for their resection make indiscriminate resection of all cystic neoplasms unjustified. Conversely, the potentially lethal nature of a cystic neoplasm that has undergone malignant transformation demands that due attention be paid to all incidentally discovered cystic pancreatic lesions. This need for attention is further highlighted by the fact that cystic pancreatic neoplasms, while constituting a minority of pancreatic tumours, are virtually curable by resection. Therefore, there is a need for an efficient process of selection of surgical candidates in whom the risks of malignant transformation of the cystic lesion are balanced by the hazards of the surgical procedure employed to resect that lesion.

Histopathological analysis of resected cystic pancreatic neoplasms has afforded tremendous insight into the characteristics of these tumours and is the most reliable means of distinguishing them. (25) Unfortunately the cost of obtaining a specimen for histopathology is the employment of the very surgical procedures that inflict morbidity on patients. Much work has been done over the past decade to replicate the accuracy of histopathology via non-invasive or minimally invasive means. Computed tomography (CT), magnetic resonance imaging (MRI), endoscopic ultrasound (EUS) and tumour marker analysis of cyst fluid aspirate have been most successful in this regard.

The sensitivity and specificity of these imaging modalities has been judged for each imaging feature that is thought to reflect a certain histopathological type of cystic neoplasm. The most widely explored imaging features in the literature are septations, mural nodules, main pancreatic duct dilation and communication of cystic lesion with pancreatic duct. Table 1 highlights some literature that explores the utility of these features in predicting for the presence of malignancy in a detected cystic pancreatic lesion.

Table 1: Utility of Cyst Features in Predicting Histology of Pancreatic Cysts (26: Sahani et al., 27 Manfredi et al.)

Imaging Feature	Statistic and Reporting Author
Cyst Septations	73.9 % Sensitivity (26) (only done for IPMNs)
	Present in 35.7 % MCNs vs. 64.3 % of SCNs (27)
Mural Nodule(s)	93 % Sensitivity, 80 % Specificity (26) (only done for IPMNs)
	Present in 94.1 % MCNs vs. 5.1 % of SCNs (27)

Cyst Wall Thickness	2- 4 mm in MCNs vs. 2 – 3 mm in SCNs (27)
Largest Cyst Diameter	40 – 73 mm in MCNs vs. 20 – 44 mm in SCNs (27)

While some of these numbers are impressive, there is a further complication: no single imaging modality has actually been found to be reliable in detecting the presence of these features. Table 2 summarises the interobserver agreement (ICC) in assessing the presence of certain features on various imaging modalities.

Table 2: Interobserver agreement in Ascertaining Presence of Cyst Features on Imaging Modalities (14: Kadiyala et al., 26: Sahani et al., 28: Jong et al., 29: Do et al.)

Imaging Feature	Interobserver Agreement	Modality	Author
Cyst Septations	0.36	MRI	(28)
	0.24	EUS	(26)
Mural Nodule(s)	0.28	CT	(29)
	0.23	MRI	(28)
	0.42	EUS	(14)
Main Pancreatic Duct Dilation	0.75	CT	(28,29)
Communication of Cystic lesion with Pancreatic Duct	0.53	MRI	(28)

Jong et al. found that the overall agreement for discriminating between mucinous and non-mucinous cysts was 0.36. (29)

The varying sensitivities, specificities, inter-observer agreements of each of these methods as well as the varying definitions used in these studies motivated the formulation of consensus guidelines on the clinical usage of diagnostic modalities to guide a multidisciplinary approach to diagnosis. The first international consensus on diagnosis and management of cystic pancreatic neoplasms was published in 2006 when a multidisciplinary panel of experts convened in Sendai, Japan. (30) This consensus guideline was revised in 2012, owing to the availability of additional literature, by another panel of experts who convened in Fukuoka, Japan. (33) Additionally, European consensus guidelines were also published in 2013. (31)

The Fukuoka consensus recommends that all patients with suspected cystic neoplasms of greater than 1 cm diameter undergo pancreatic-protocol CT or gadolinium-enhanced MRI with magnetic resonance cholangiopancreatography (MRCP). Asymptomatic cysts that are < 1 cm size have rarely been seen to harbour malignant elements and therefore no further workup is needed.

Management is currently guided by the findings of

- cyst size > 3 cm
- "Worrisome features" (thickened cyst walls, main pancreatic duct (MPD), calibre of 5 – 9 mm, an abrupt change in MPD calibre with distal pancreatic atrophy, non-enhanced mural nodules, lymphadenopathy)

- "High risk stigmata" (obstructive jaundice with cystic lesion of pancreatic head, enhanced solid component, MPD size > 10 mm).

If none of these is found, the patient is followed by serial imaging. In this case, additional information regarding the risk of malignancy is afforded by determining the type of cystic mass. Clues towards determining type of cystic mass are elicited by examining morphology, multiplicity, location, and communication with MPD on imaging, as well as CEA and cytology of fine needle aspirates. CEA helps distinguish mucinous from non-mucinous cysts but does not discriminate malignant from benign lesions. Dysplastic cells seen on cytology give clues towards malignancy but this is only recommended in centres with expertise.

Having identified these features in the consensus, the surgical community now has a guided pathway (32) to organize multidisciplinary efforts in diagnosis. The guided pathway allows utilization of radiological and biochemical features with aforementioned sensitivities and specificities in a manner that maximally allows differentiation between malignant and non-malignant cysts. Goh et al. found the positive and negative predictive value of a patient being classified as high risk for malignancy as per the revised 2012 Fukuoka guidelines to be 88% and 92.5% respectively. (33)

While differentiating between malignant and non-malignant cysts is a key question in clinical management, the consensus guidelines still miss out on the primary instinct that clinicians adopt in the management of any disease which is to assess the natural history of the disease. The consensus guidelines do not primarily address the matter of predicting for the histological subtype of the pancreatic cyst while this knowledge is still expected to be employed in clinical judgement. It is difficult to assess for malignancy without trying to gain an idea of the

histological type of the lesion since features that otherwise predict for malignancy are sometimes also found in SCNs which are less predisposed to malignant transformation.

Furthermore, there is a net loss of sensitivity over the steps of the pathway as features are tested sequentially rather than simultaneously. Demographic characteristics that do show a differential distribution over tumour categories, such as age and gender, are not directly included in the diagnostic pathway yet are expected to be employed in clinical judgement while diagnosing the cyst. All-in-all, significant expertise in clinical judgement is required to overcome the subjective elements of the diagnostic process namely the poor inter-observer agreement in detecting these features as well as the need to judge the underlying histology of the cyst. Such expert clinical judgement is time-consuming to develop and potentially restricted to high volume centres. Kleeff et al. found that 10% of patients were misclassified preoperatively and could potentially have avoided surgery. (23)

1.3 Study Goals and Objectives

The goal of this study is to provide a tool that will simplify the non-invasive detection of the histology of pancreatic cystic lesions for general and pancreatic surgeons, a process that otherwise requires expert clinical judgement, by harnessing elements of clinical judgement in a statistically guided framework.

The objective of the study is to apply a statistical classification method to case-specific demographic and radiological information. The statistical method must be such that it provides an explicable rule by which surgeons may be able to discriminate between Serous Cystic

Neoplasms, Mucinous cystic neoplasms and Intraductal Papillary Mucinous Neoplasms with test characteristics that make it equivalent to, if not better than, experienced surgical opinion.

2. Methodology

2.1 Patient Selection

In this cross-sectional study, inpatients and outpatients with a confirmed histopathological diagnosis of either SCN, MCN or IPMN since year 1990 to date of age greater than 18 years were queried from a prospectively maintained database at the Johns Hopkins Medical Institutions dedicated to cystic pancreatic pathologies. Patients of age 18 and below were excluded since the distribution of cystic pathologies seen in that age group is not of primary interest in this study. The following data were either extracted from this database or from the Johns Hopkins electronic medical record regarding the patients that resulted from the query and used as variables for subsequent development of the model:

1) Age
2) Gender
3) Ethnicity
4) Symptoms and signs, if any, at presentation
    a) Pruritus
    b) Jaundice defined as yellow discolouration of eyes and/or skin either reported by the patient or noted by the examining clinician
    c) Weight Loss as defined as a loss of 10 lbs or subjective feeling of loosening of clothing

d) Abdominal Pain

e) Past History of any malignancy

f) Past History of diabetes mellitus

5) Family History of pancreatic malignancy, cystic or otherwise

6) The following CT, MRI and endoscopic ultrasound image findings as reported by the clinical radiologist seeing the patient regarding the largest cystic lesion present in the patient at the time of presentation

a) Cyst Largest Diameter

b) Location of cyst

c) Presence of a thick cyst wall, where thick is defined as greater than 2 mm

d) Presence of calcification(s) in the cyst

e) Presence of septations(s) in the cyst

f) Presence of mural nodule(s) in the cyst

g) Presence of main pancreatic duct dilation in the pancreas

h) Presence of common bile duct dilation

This yielded 900 cases with 1804 total images.

Following this inclusion, patients placed on radiological surveillance upon diagnosis were excluded from the results since they did not have a definite histological confirmation of diagnosis. Furthermore, repeat images of the cystic lesion of patients who were placed on surveillance prior to obtaining a definite histological confirmation of the diagnosis were discarded. This is because the classification tree method is currently not implemented for

longitudinal or panel data (see Model Development for details). Patient with multiple cystic lesions were only followed up till the surgical excision of the largest cyst. Thereafter, the final dataset contained 711 cases with 980 total images.

Patients who are found to have more than one category of cystic pancreatic pathology were not excluded as the features of each cyst were studied and applied to the model individually. This applies to patients in whom the discovered cysts are not of the histological category of interest in this study. Figure 1 summarises the patient selection processes.

2.2 Missing Data

Data on the subjective sensation of pruritus, family history of pancreatic cancer and past history of malignancy was not available for all cases from the electronic medical record. Furthermore, certain radiological data were not available in all imaging reports in the electronic medical record. For complete and minimally biased analysis, data was imputed under a mathematical framework called multiple imputation by chained equations, a technique now recommended as standard practice. By this framework, each variable was imputed by borrowing information from the remaining variables called multiple imputation by chained equations. (34) In this technique, missing data in each variable is imputed via a regression on the remaining variables that were highly correlated with variable containing missing data. (35)

2.3 Model development

The model that appropriately addresses the problem laid out previously must have the following properties:

- It must be able to address a trichotomous outcome (i.e. prediction of either SCA, MCN or IPMN) where outcomes are considered mutually exclusive
- It must be explicable for example by expressing the effect of each variable towards the prediction of any particular outcome

  This is essential for its acceptability in routine clinical practice
- It should ideally account for uncertainly in predictors (independent variables)

  Since all imaging modalities are not proved to have adequate reliability in detecting features the model should account for uncertainly in the data
- It should ideally account for longitudinal data since suspected cystic pancreatic neoplasms may be placed on surveillance if immediate surgical intervention is not desired

Explicable statistical methods that account for trichotomous outcomes are multinomial logistic regression and a tree-based classification algorithm known as conditional inference tree which have been described previously. (36,37)

Briefly, the multinomial logistic regression model is exactly the same as the logistic regression model in that one of the outcome categories is considered baseline. The regression coefficients therefore describe the effects of the independent variables on the probability of predicting for the alternate outcome as opposed to the baseline outcome on a log scale. The difference is that the multinomial logistic model is capable of running two or more regressions against the baseline outcome. These regressions inform each other through the constraint that their predicted probabilities must sum to one.

The multinomial logistic regression may be executed under a Bayesian framework to account for uncertainty in the predictors as described by Ntzoufras (38). Non-informative prior distributions is assigned to the regression coefficients. The standard deviations of the prior distributions of categorical variables may be linked by hyper-priors. (39) This is done because the presence of one level of a categorical variable is in fact informative of the absence of the rest. Finally, a random effect variable may be introduced to account for multiple surveillance images taken before deciding to operate on a patient.

The Conditional Inference Tree also has its roots in statistical inference methods as opposed to traditional tree based methods. this algorithm produced a decision tree by examining variables of the model one-by-one to choose the one that gave the greatest classification power. Classification power is defined in this methodology on the basis of information theory which searches for the minimum number of questions needed to be asked in order to arrive at the desired answer. In this case, the algorithm searched for the least number of questions that need to be answered about the cyst in order to determine its histological subtype.

After selecting a variable, the algorithm examined splits (for categorical variables) or cut-off points (for continuous variables) in the selected variable once again to select one that afforded the greatest classification power. Both variable and split selections were guided by statistical criteria based on p-values. This process was repeated to produce steps of a decision tree until the statistical criteria could not be satisfied further. Variables that had already been selected at a higher level in the decision tree could be reconsidered for a lower level in the decision tree.

Unfortunately there does not currently exist and implementation for classification trees under the Bayesian framework. Moreover, the conditional inference tree does not yet support longitudinal data. It would therefore seem that Bayesian multinomial logistic regression would be the most appropriate tool for this problem. However, this methodology requires a few important considerations in order for its application to be legitimate. Agresti states that, as a rule of thumb, the dataset must contain 10 - 25 cases for each independent variable included in the model. The introduction of priors (and hyper-priors) demands that this principle be applied conservatively (i.e. 10). In this case, the question arises as to which of the independent variables under consideration can be included in this model.

One way is to train multiple models by forward or backward selection. This is not feasible since the Bayesian framework is implemented computationally, and not mathematically, via a set of sampling methods known as Markov Chain Monte Carlo methods. The properties of these methods, demand that computations with large number of iterations be run in order to produce results that are not auto-correlated and reflect the true posterior distribution of the coefficients (see Kruschke (39) for details). With the aforementioned sample size, such a computation can take days to complete even on a computational cluster. Therefore, more efficient method is needed to select independent variables of interest.

The properties of the Conditional Inference Tree make it suitable for this task. Variables with poor reliability will not be very good at predicting for the outcome. These variable will not make it through the variable selection process of the Conditional Inference Tree. Therefore, at the expense of the ability to examine the effects of unreliability and longitudinal data on prediction, the Conditional Inference Tree was chosen. (This is why pre-operative surveillance images were

excluded in the selection process) The intention was to develop the rudiments of a tool incorporating the bare essentials for predicting for cystic pancreatic neoplasms histology and examining the performance yielded therein.

2.4 Model Evaluation

The resulting decision tree was then evaluated for its ability to accurately classify cases into the correct histological subtype of cystic pancreatic neoplasm by constructing a receiver operating curve surface (ROCS), a methodology put forward by Mossman. This surface is equivalent to the receiver operating curve except it is designed to evaluate the performance of three outcomes groups. The volume under the ROCS (VUS) was used as a summary measure of assessment of classification. This is equivalent to the area under the receiver operating curve (AUC) except, once again, it is designed to evaluate three outcome groups. (40) Furthermore, the models' ability to distinguish between SCN vs. non-SCN category of cystic pancreatic neoplasm was examined considering that misclassification in this regard had the greatest cost.

2.5 Sensitivity Analysis

To understand the effects of uncertainty in the data on the performance of the decision tree, five sets of data were imputed using the imputation process described above. Doing so had the effect of producing datasets that preserved the properties of the original dataset such as the proportion of levels in categorical variables yet introducing a "jitter" in the dataset making each imputed dataset slightly different. The decision tree was then trained on the first dataset and tested against each of the other four datasets in a manner similar to Carpenter at al. (41) The four resulting VUS were compared against each other.

Subsequently, the decision tree was evaluated by cross-validating bootstrapped datasets (42): three-fourths of the dataset was randomly selected (with replacement) to create a training set on which the decision tree was trained following which the remaining one-fourth was used to test the trained decision tree. This process was repeated 10,000 times to produce a distribution of VUS estimates. This bootstrap method provided a more rigorous method of sensitivity analysis whereby the properties of the original dataset were not preserved allowed for examination of performance under "extreme" conditions.

2.6 Software

The entirety of the study was conducted in the *R* statistical programming language. (43) Missing data imputation was performed using the *mice* package (35). The conditional inference tree was developed using the *partykit* package (37). The 3-way ROCS and the volume under the ROCS was computed by programming Mossman's mathematical logic into *R* code. The trained decision tree was programmed into a simple interface requiring the elements needed to make a decision using the *shiny* package. (44)

3. Results

3.1 Patient Characteristics

Table 3 summarises the characteristics of patients according to diagnosed cyst type. The dataset consisted of 711 patients, comprising 422 IPMNs (60.9%), 219 SCNs (27.4%) and 71 MCNs (11.6%). The mean age (SD) of the patients was 64.1 ± 13.6 years. Fifty-nine percent of the patients were female. Eighty-two percent of the population were Whites with the

remainder comprising of 9.6 %, 3.2 %, 4.5 % African Americans, Asians and other respectively. Sixteen percent of the patients had diabetes. Eighteen percent of the patients had a past history of malignancy. Fourteen percent of the patients had a family history of malignancy. The mean (SD) largest diameter of cysts was 3.73 ± 2.50 mm on CT. Forty percent of cysts were location in the tail region, 39.8 % in the head region and the remainder in the neck region. Cyst characteristics have been reported for CT images since these images have the least amount of missing data.

Table 3: Patient and Cyst Characteristics

Characteristic	All Samples (n = 711)	IPMN (n = 421)	SCN (n = 219)	MCN (n = 71)
Patient Characteristics, n unless otherwise specified				
Female Sex	422	203	150	69
Ethnicity				
White	588	372	169	47
African American	68	24	29	15
Asian	23	8	11	4
Other	32	17	10	6
Age at	64.1 ±13.6	68.4 ±11.4	60.2 ±14.4	50.3 ± 11.7

Surgery (mean ± SD)				
Presence of Abdominal Pain	289	190	75	24
Pruritus (n = 530)ξ	26	19	7	0
Weight Loss	107	96	10	1
Jaundice	66	61	5	0
Diabetes (n = 705)ξ	112	74	29	9
Nausea and Vomiting (n = 667)ξ	45	32	3	10
History of Past Malignancy	112	56	53	3

(n = 628)ξ												
Family History of Pancreatic Cancer (n = 446)ξ	64			44			14			6		
Cyst Characteristics, n unless otherwise specified												
	CT	MRI	EUS	CT	MRI	EUS	CT	MRI	EUS	CT	MRI	EUS
Mean Cyst Diameter, mm	3.73	3.60	2.86	3.06	2.60	2.52	4.37	4.56	3.47	4.69	4.16	3.23
Location of Cyst in Pancreas												
Head Region	259	33	92	186	22	69	72	10	23	1	1	0
Neck Region	47	5	7	27	3	5	16	1	2	4	1	0
Tail Region	265	57	54	84	14	42	115	16	38	66	27	4

Presence of Cyst Septations	211	30	86	108	18	77	83	0	0	20	12	2
Presence of Calcifications in Cyst	73	6	10	22	1	1	38	1	2	13	4	1
Presence of Mural Nodule(s) in Cyst	115	10	49	55	4	37	54	4	12	6	2	0
Presence of Thick Cyst Wall in Cyst	28	2	13	20	1	13	3	0	0	5	1	0
Presence of Main Pancreatic Duct Dilation on imaging	228	28	86	198	22	84	27	4	2	3	2	0

Presence of Common Bile Duct Dilation on Imaging	71	7	0	67	4	0	4	1	0	0	2	0

ξ data reported is that which was available from the electronic medical record

η data reported from CT only

3.2 Decision Tree

Figure 2 shows the decision tree as developed by the conditional inference tree algorithm. The tree branches at variables which were statistically significant for providing classification power; these variables are as follows:

1. Calcification of the cyst (CALCI)
2. Common Bile Duct Dilation (CBD)
3. Location of the tumour within the pancreas (LOCATION)
4. Head Region
5. Neck Region
6. Tail Region
7. Age of the patient (AGE)
8. Largest diameter of the cyst (LARGEST)
9. Main Pancreatic Duct Dilation (MPD)
10. Presence of a thick cyst wall (WALL)

11. History of Past Malignancy in the patient (pancreatic or other) (PAST.MALIG)
12. Gender of the patient (GENDER)

The words in parenthesis correspond to the abbreviations used at the branches of the decision tree. The type of imaging modality used to detect the cystic lesion was not significant in the tree building algorithm along with:

1. Ethnicity
2. History of weight loss
3. History of abdominal pain
4. History of jaundice
5. History of pruritus
6. History of nausea and/or vomiting
7. Family history of pancreatic cancer
8. History of diabetes
9. Presence of septations in the cyst
10. Presence of mural nodule(s) in the cyst

3.3 Evaluation

Figure 3 shows the receiver operating curve surface for the decision tree as trained against the first imputed dataset and tested against the second. The volume under the surface (meaning the probability of correctly classifying SCNs, MCNs and IPMNs) was 77.5% ($p < 0.001$). The remaining 23% that was not correctly classified comprised of 36.2% of SCNs that were misclassified as IPMNs or MCNs and 15.9% of non-SCN cases that were misclassified as SCNs.

Table 2 shows VUSs obtained by testing the decision tree against all imputed data sets.

Test Set	**VUS**
Imputed Set 2	77.5
Imputed Set 3	77.4
Imputed Set 4	72.4
Imputed Set 5	76.8

Table 4: Volumes Under Surface for testing against imputed data sets

Figure 4 shows the distribution of bootstrapped VUS values. The mean VUS value was 66% with a 95% confidence interval of 54.5% - 78.5%.

3.4 Clinical Application

Figure 5 shows a screenshot of 'shiny' interface to the trained decision tree. It displays the most likely outcome as well as displays a bar chart of the probabilities of the other possibilities.

4. Discussion

Adequate diagnosis of cystic pancreatic lesions as yet remains a matter of expert opinion especially because of the multidisciplinary approach required and because a significant amount of experience is necessary to overcome the subjective elements in the process. The decision tree developed in this paper provides a solid foundation for capturing this complex process to harness it into a simple interface. Such an interface can imbue efficiency into the diagnosis process on a larger scale that is not limited by the time it takes to develop the clinical acumen that traditionally has allowed such efficiency. It can do so with an accuracy that is equivalent to current expert clinical practice.

The decision tree has simplified the diagnostic process by narrowing it to the most statistically significant variables necessary to achieve maximum diagnostic accuracy. Furthermore, the interface to the decision tree provides a simple checklist to coordinate the data collection from different modalities involved.

Many a times, the pancreas is peppered with cysts and the surgeon may only want to resect the one(s) that appear most threatening in the interest of preserving some pancreatic function. The decision tree has been developed with the intention of focusing on a single cyst thereby supporting considerations in resection and surveillance strategies to be made at an individual cyst level.

The decision tree has been developed with a basic clinical question in mind: what is the histology of the cystic lesion in question? Answering this question has two major utilities: first, the clinician can incorporate biological knowledge regarding the natural history and potential

for malignancy of the cystic lesion to assist decision to resect versus surveillance; second, the clinician can also use to decision tree to “tag” cystic lesions in patients who have been placed on surveillance in order to follow them in the long run to study the nature and appearance of malignant features. The latter utility offers the advantage of more detailed understanding of the individual natural histories of each of the cystic pancreatic neoplasms, which in turn can aid tailoring of resection and surveillance strategies specific to each subtype.

Moreover, this foundation can be built upon to push the accuracy of diagnosis higher (we use the term accuracy for a lack of equivalent terms to sensitivity and specificity in the 3 outcome case). Recent studies have highlighted the potential value of molecular markers (45) in providing this added accuracy and such data could be incorporated into the decision tree. Other, though lower yield, additions entail further exploration of biochemical markers in cyst fluid. Ryu et al., in a cohort of 56 cystic fluid aspirates, identified that 86.7% of the mucinous cystic tumours (MCAs and IPMNs combined) displayed CEA levels > 400 ng/dL whereas the highest CEA level among 8 SCNs was 57 ng/dL. They concluded that a CEA threshold of 467 ng/dL had a sensitivity of 87% and a specificity of 98% for distinguishing between mucinous and cysts of any other kind. CA 19-9, on the other hand, did not reveal any discriminatory power between cysts. (46) Van der Waaij et al., in their pooled analysis of 12 studies, propose a lower CEA cut-off of < 5 ng/mL that distinguishes between mucinous and serous lesions with 54% sensitivity and 94% specificity. They also propose a CA 19-9 cut-off of 37 ng/mL to exclude SCNs with sensitivity 19% and specificity 98%. (47)

Brugge et al. report sensitivities and specificities from a compendium of studies that attempted to diagnose cystic lesions with CEA with cut-off > 400 ng/mL for mucinous lesions and < 4

ng/mL for serous lesions. 14 serous cystadenomas were diagnosed resulting in 54% sensitivity and 77% specificity and 17 MCAs were diagnosed resulting in 13% sensitivity and 75% specificity. (12)

The decision tree also needs to be extended to cover peculiar situations that were not specifically addressed in this study. For example, some patients present with isolated dilation of main pancreatic duct as the only abnormal finding which may actually be a sign of an IPMN. As detailed in the Introduction section, SCNs also have an oligocystic variant that has no septations or locules and are typically positioned in the pancreas in a manner that make them hard to distinguish from MCNs. Addressing these variations in presentation of cysts may be key to reducing the misclassification rate of the decision tree.

4.1 Limitations

Should there ever be discovered a single modality that diagnoses the category of cystic pancreatic neoplasm with very high accuracy (> 80 % in the least), then such a modality would obviate the need for this decision tree and modalities employed therein.

While the decision tree is robust in the face of minor changes to the data set, it is very sensitive to the proportion of the tumours in the population being predicted for. This is reflected by the bootstrapped values of VUS shown in the previous section. When the proportion of tumours in the training and test sets was similar, the performance of prediction exceeded 75%. Conversely, when the proportion of tumours in the training and test sets was not similar, then the performance of prediction deteriorates going as low as 38%. Therefore, if the decision tree is

used to predict for a patient from a population wherein the proportion of tumours does not match that of the training set, then an erroneous prediction is to be expected.

Another factor to consider is that the decision tree was developed on cases with confirmed pathological diagnosis meaning that all cystic lesions in these cases were considered "concerning" enough to necessitate surgical excision. This biases our results in that the decision is only applicable to patients with "concerning" cystic lesions in the pancreas.

One factor that is fairly arduous to address is the fact that the decision tree was trained on cases referred to a tertiary care referral centre that many not be representative of the true proportion of tumours in the general population. Concordantly, this decision tree cannot be used in community clinics or in specialties outside of general surgery. One may expand the dataset by repeating the process in a cross-institutional dataset, but the fact remains that large scale and long-term tracking of cystic pancreatic neoplasms is only currently performed at tertiary referral centres.

The decision tree assumes that when a particular feature regarding the cyst is reported, for example the presence of cyst calcification, then that features is either definitively present or absent. However, in reality poor inter-rater reliability has been reported in detecting certain cyst features via imaging modalities. [reference] A Bayesian model may be more appropriate in this scenario as uncertainty may lead to different conclusions in predicting for the histology of cystic pancreatic neoplasms.

The behaviour of the decision tree is such that it, whenever it misclassifies a case, it tends to misclassify cases in favour of those tumours that are historically known to be more malignant

(i.e. more SCNs were misclassified as IPMNs and MCNs than vice versa). This behaviour may in fact be an indication that decision tree is capturing the decision-making trends of the surgical consultants over the years. Indeed the Fukuoka consensus guidelines were designed to err on the side of caution in a similar sense and post-hoc studies that have attempted to assess the performance of the Fukuoka consensus guidelines have confirmed as such. This behaviour is, therefore, concerning that the decision tree is nothing more than a mathematical representation of clinician judgement. Since clinician judgment has been primary oriented towards patient welfare and early removal of a potentially curable malignancy, it deviates from the primary goal of this study, which was to develop a tool that non-invasively elucidates the underlying histology of the cystic pancreatic lesion. Future revisions to this statistical model will need to eliminate the effect of the clinician judgment.

Following the same concern, an important consideration is to employ all diagnostic modalities with equal frequency as opposed to current clinician practice, which is to do a EUS with FNA most if diagnosis of the cystic lesion on CT and/or MRI remains ambiguous.

Notwithstanding these limitations, an alternate approach has been detailed in the Methodology section that is capable of addressing them. The multinomial logistic regression model, computed under the Bayesian framework, can now be used with the variables selected by the decision tree. Under these variables, the regression model can be developed with the confidence of equal, if not higher, predictive power while examining the effects of uncertainty and longitudinal data from pre-operative surveillance images. Furthermore, latent variables can be incorporated to account for surveillance cases that do not yet have confirmed pathological diagnosis to increase the generalizability of the tool to a broader range of patients.

5. Conclusion

In conclusion, this study has produced a clinically applicable tool that can be used alongside the current Fukuoka guidelines to enhance decision making ability. It can do so for a broad population of surgeons that have yet to be galvanised by the experience it takes to master the practice of the Fukuoka guidelines. It provides the base for further work that may enhance predictive power especially with regard to patients who do not require surgery.

6. Figures

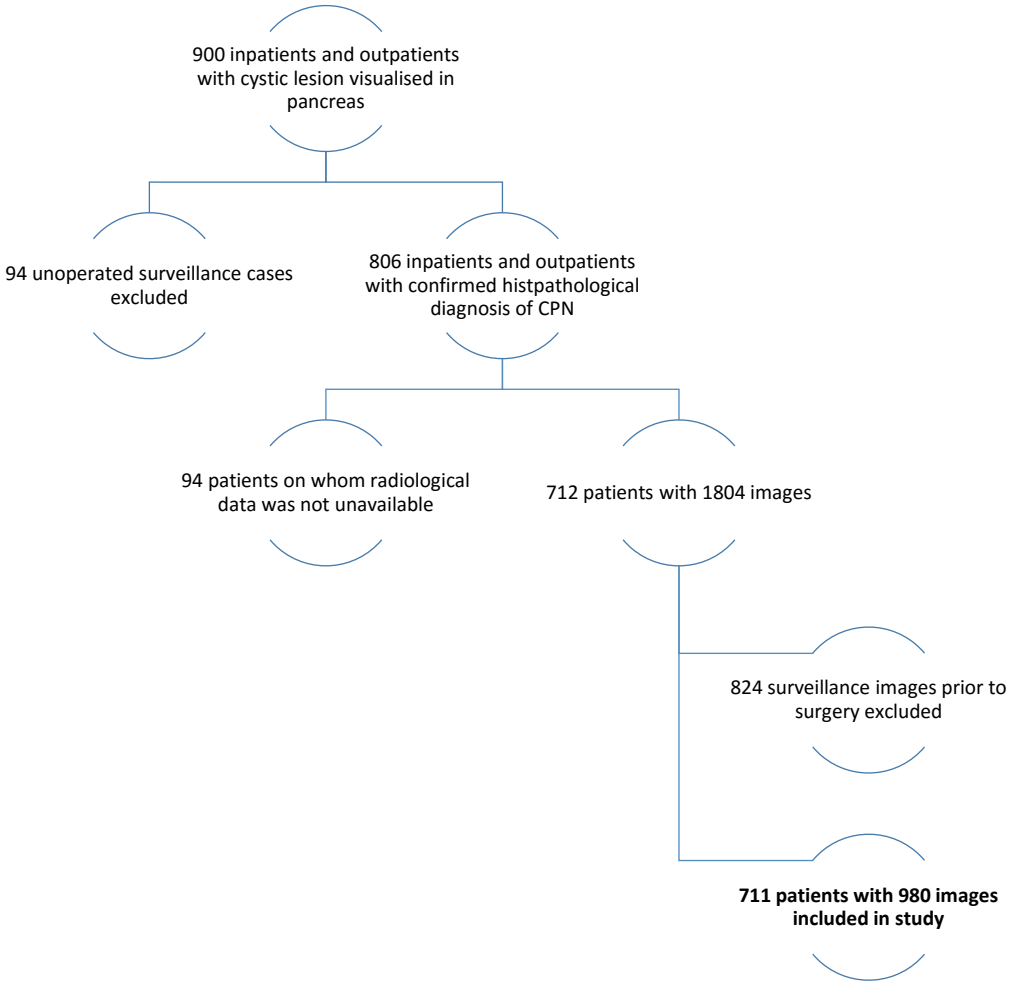

Figure 1: Patient Selection

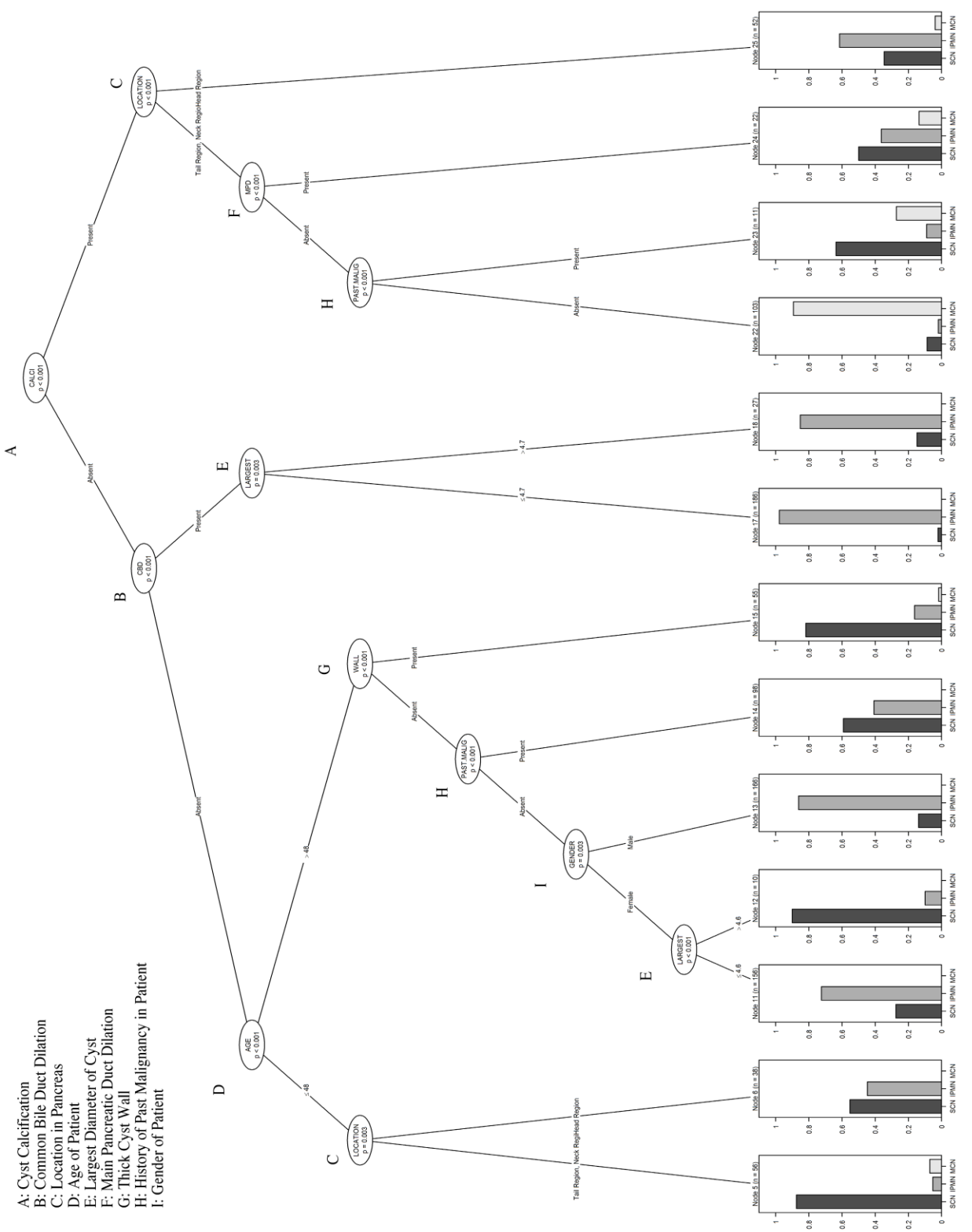

Figure 2: Decision Tree as developed by the Conditional Inference Tree algorithm

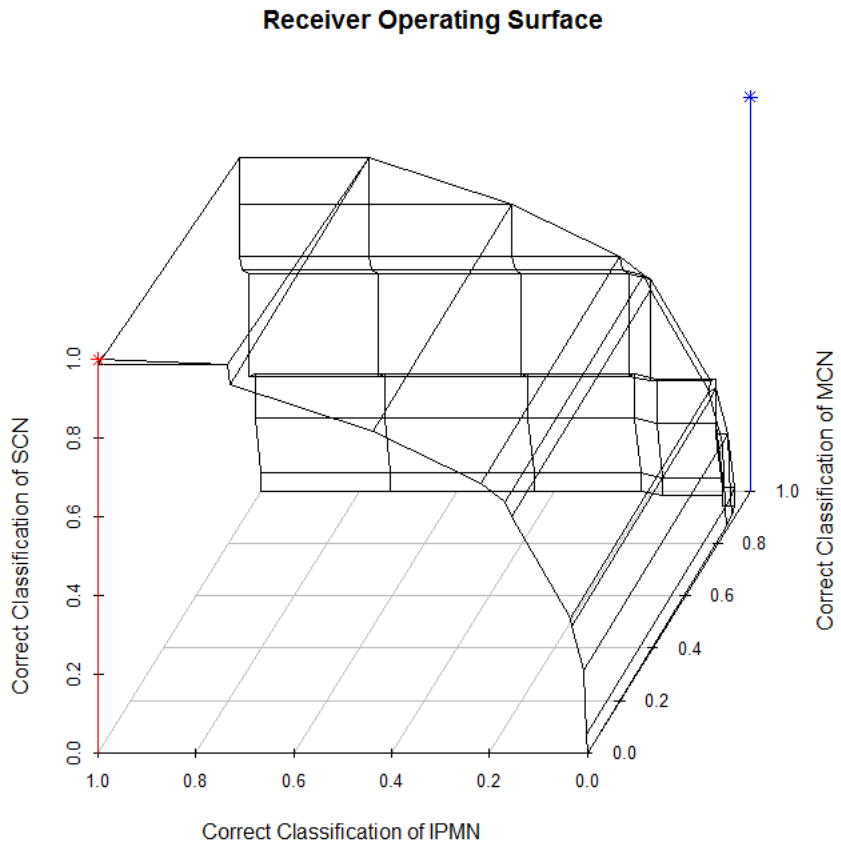

Figure 3: Receiver Operating Curve Surface (the surface begins at the red star; a perfect surface would look like a cube with the outer most corner touching the blue star

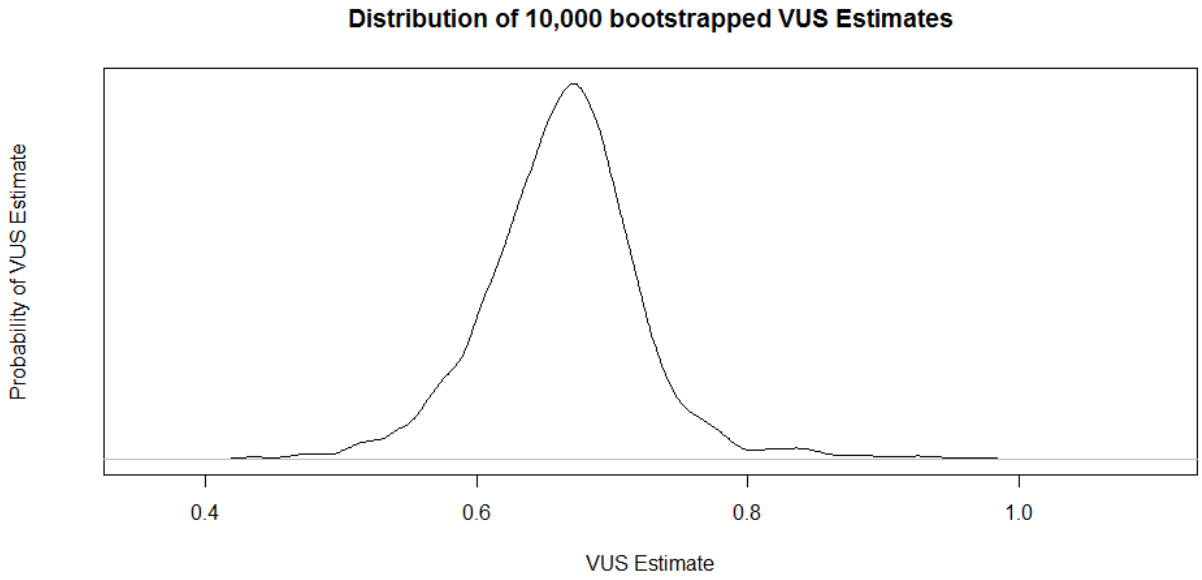

Figure 4: Empirical Distribution of VUS (a value of 0.25 is synonymous with a test that offers no classification power)

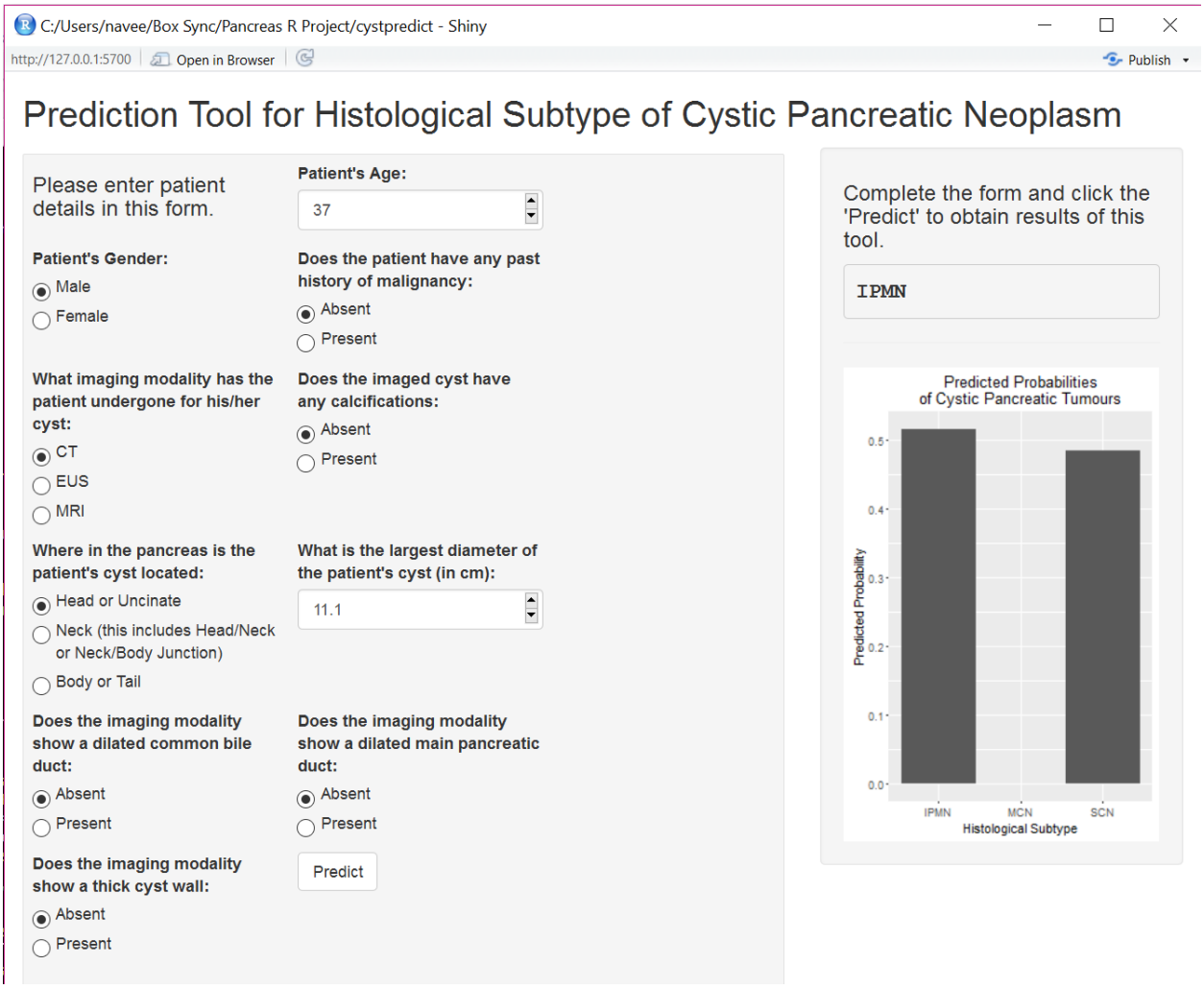

Figure 5: A simple interface developed using the 'Shiny' framework for predicting cyst histology

REFERENCES

1. Hruban RH, Boffetta P, Hiraoka N, Iacobuzio-Donahue C, Kato Y, Kern SE, et al. Tumours of the Exocrine Pancreas. WHO Classif Tumours Dig Syst. 2010;281–337.

2. Krušlin B, Zovak M, Doko M, Belicza M. Serous oligocystic and ill-demarcated adenoma of the pancreas. Virchows Arch. 2002;440(4):441–2.

3. Kimura W, Nagai H, Kuroda a, Muto T, Esaki Y. Analysis of small cystic lesions of the pancreas. Int J Pancreatol. 1995;18(3):197–206.

4. Pitman MB, Centeno BA, Ali SZ, Genevay M, Stelow E, Mino-Kenudson M, et al. Standardized terminology and nomenclature for pancreatobiliary cytology: The Papanicolaou Society of Cytopathology guidelines. Diagn Cytopathol. 2014;42(4):338–50.

5. Hopwood D. Histological Typing of Tumours of the Exocrine Pancreas. J Clin Pathol. 1996;49(9):780.

6. Azzopardi N. Cystic lesions of the pancreas. 2014;26(02):58–63.

7. Klöppel G, Basturk O, Schlitter AM, Konukiewitz B, Esposito I. Intraductal neoplasms of the pancreas. Semin Diagn Pathol [Internet]. W.B. Saunders; 2014 Dec [cited 2015 Mar 31];31(6):452–66. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84916628437&partnerID=tZOtx3y1

8. Cooper CL, O'Toole S a, Kench JG. Classification, morphology and molecular pathology of premalignant lesions of the pancreas. Pathology [Internet]. 2013;45(3):286–304. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23442735

9. Sahani D V, Kadavigere R, Saokar A, Fernandez-del Castillo C, Brugge WR, Hahn PF. Cystic pancreatic lesions: a simple imaging-based classification system for guiding management. Radiographics. 2005;25(6):1471–84.

10. Cho HW, Choi JY, Kim MJ, Park MS, Lim JS, Yong EC, et al. Pancreatic tumors: Emphasis on CT findings and

pathologic classification. Korean J Radiol. 2011;12(6):731–9.

11. Pedrosa I, Boparai D. Imaging considerations in intraductal papillary mucinous neoplasms of the pancreas. World J Gastrointest Surg. 2010;2(10):324–30.

12. Brugge WR. Evaluation of pancreatic cystic lesions with EUS. Gastrointest Endosc. 2004;59(6):698–707.

13. Gress F, Gottlieb K, Cummings O, Sherman S, Lehman G. Endoscopic ultrasound characteristics of mucinous cystic neoplasms of the pancreas. Am J Gastroenterol. 2000;95(4):961–5.

14. Kadiyala V, Lee L. Endosonography in the diagnosis and management of pancreatic cysts. World J Gastrointest Endosc [Internet]. 2015;7(3):213. Available from: http://www.wjgnet.com/1948-5190/full/v7/i3/213.htm

15. Scoazec JY, Vullierme MP, Barthet M, Gonzalez JM, Sauvanet a. Cystic and ductal tumors of the pancreas: diagnosis and management. J Visc Surg [Internet]. 2013;150(2):69–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23518192

16. Linder JD, Geenen JE, Catalano MF. Cyst fluid analysis obtained by EUS-guided FNA in the evaluation of discrete cystic neoplasms of the pancreas : a prospective single-center experience. 2006;64(5):697–702.

17. Brugge WR, Lewandrowski K, Lee-Lewandrowski E, Centeno B a., Szydlo T, Regan S, et al. Diagnosis of Pancreatic Cystic Neoplasms: A Report of the Cooperative Pancreatic Cyst Study. Gastroenterology. 2004;126(5):1330–6.

18. Kalb B, Sarmiento JM, Kooby D a, Adsay NV, Martin DR. MR imaging of cystic lesions of the pancreas. Radiographics. 2009;29(6):1749–65.

19. Wang Y, Miller FH, Chen ZE, Merrick L, Mortele KJ, Hoff FL, et al. Diffusion-weighted MR imaging of solid and cystic lesions of the pancreas. Radiographics. 2011;31(3):E13–30.

20. Barral M, Soyer P, Dohan A, Laurent V, Hoeffel C, Fishman EK, et al. Magnetic resonance imaging of cystic pancreatic lesions in adults: An update in current diagnostic features and management. Abdom Imaging.

2014;39(1):48–65.

21. Gaujoux S, Brennan MF, Gonen M, D'Angelica MI, DeMatteo R, Fong Y, et al. Cystic lesions of the pancreas: Changes in the presentation and management of 1,424 patients at a single institution over a 15-year time period. J Am Coll Surg [Internet]. Elsevier Inc.; 2011;212(4):590–600. Available from: http://dx.doi.org/10.1016/j.jamcollsurg.2011.01.016

22. Fernández-del Castillo C, Targarona J, Thayer SP, Rattner DW, Brugge WR, Warshaw AL. Incidental pancreatic cysts: clinicopathologic characteristics and comparison with symptomatic patients. Arch Surg. 2003;138(4):427–423; discussion 433–4.

23. Kleeff J, Michalski C, Kong B, Erkan M, Roth S, Siveke J, et al. Surgery for Cystic Pancreatic Lesions in the Post-Sendai Era : A Single Institution Experience. Hindawi Publishing Corporation; 2015;2015.

24. Schmidt CM, Powell ES, Yiannoutsos CT, Howard TJ, Wiebke EA, Wiesenauer CA, et al. Pancreaticoduodenectomy: a 20-year experience in 516 patients. Arch Surg. 2004;139(7):718–25; discussion 725–7.

25. Adsay NV. Cystic lesions of the pancreas. Mod Pathol. 2007;20:S71–93.

26. Sahani D V., Kambadakone A, MacAri M, Takahashi N, Chari S, Fernandez-Del Castillo C. Diagnosis and management of cystic pancreatic lesions. Am J Roentgenol. 2013;200(February):343–54.

27. Manfredi R, Ventriglia A, Mantovani W, Mehrabi S, Boninsegna E, Zamboni G, et al. Mucinous cystic neoplasms and serous cystadenomas arising in the body-tail of the pancreas: MR imaging characterization. Eur Radiol [Internet]. 2014; Available from: http://link.springer.com/10.1007/s00330-014-3493-2

28. de Jong K, Nio CY, Mearadji B, Phoa SS, Engelbrecht MR, Dijkgraaf MG, et al. Disappointing Interobserver Agreement Among Radiologists for a Classifying Diagnosis of Pancreatic Cysts Using Magnetic Resonance Imaging. Pancreas. 2012;41(2):278–82.

29. Do RKG, Katz SS, Gollub MJ, Li J, LaFemina J, Zabor EC, et al. Interobserver Agreement for Detection of

Malignant Features of Intraductal Papillary Mucinous Neoplasms of the Pancreas on MDCT. Am J Roentgenol [Internet]. 2014;203(5):973–9. Available from: http://www.ajronline.org/doi/abs/10.2214/AJR.13.11490

30. Tanaka M, Chari S, Adsay V, Fernandez-del Castillo C, Falconi M, Shimizu M, et al. International consensus guidelines for management of intraductal papillary mucinous neoplasms and mucinous cystic neoplasms of the pancreas. Pancreatology. 2006;6(1-2):17–32.

31. Del Chiaro M, Verbeke C, Salvia R, Klöppel G, Werner J, McKay C, et al. European experts consensus statement on cystic tumours of the pancreas. Dig Liver Dis. 2013;45(9):703–11.

32. Lennon AM, Wolfgang C. Cystic Neoplasms of the Pancreas. J Gastrointest Surg. 2013;17:645–53.

33. Goh BKP, Tan DMY, Thng C-H, Lee S-Y, Low ASC, Chan C-Y, et al. Are the Sendai and Fukuoka Consensus Guidelines for Cystic Mucinous Neoplasms of the Pancreas Useful in the Initial Triage of all Suspected Pancreatic Cystic Neoplasms? A Single-Institution Experience with 317 Surgically-Treated Patients. Ann Surg Oncol [Internet]. 2014;21(6):1919–26. Available from: http://link.springer.com/10.1245/s10434-014-3501-4\npapers3://publication/doi/10.1245/s10434-014-3501-4

34. Newgard CD, Lewis RJ. Missing Data How to Best Account for What Is Not Known. JAMA (Journal Am Med Assoc. 2015;314(9):940–1.

35. van Buuren S, Groothuis-Oudshoorn K. {mice}: Multivariate Imputation by Chained Equations in R. J Stat Softw [Internet]. 2011;45(3):1–67. Available from: http://www.jstatsoft.org/v45/i03/

36. Agresti A. Analysis of Ordinal Categorical Data - Examples of Using R for Modeling Ordinal Data Summary of R (and S-Plus). Analysis. 2010;2010:30.

37. Hothorn T, Hornik K, Zeileis A. ctree: Conditional Inference Trees. CranAtR-ProjectOrg. 2006;

38. Ntzoufras I. Bayesian Modeling Using WinBUGS. Comput Stat [Internet]. 2011;698:592. Available from: http://books.google.com/books?id=1yZnwnOxXsEC

39. Kruschke JK. Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Eur J Psychol. 2011;7(4):1–187.

40. Mossman D. Three-way ROCs. Med Decis Making. 1999;19(1):78–89.

41. Carpenter JR, Kenward MG, White IR. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. Stat Methods Med Res [Internet]. 2007;16(3):259–75. Available from: http://ezproxy.ecu.edu.au/login?url=http://search.proquest.com/docview/217686976?accountid=10675

42. James G, Witten D, Hastie T, Tibishirani R. An Introduction to Statistical Learning. In: Springer Texts in Statistics. Sprin; 2013. p. 426.

43. R Core Development Team. R: a language and environment for statistical computing, 3.2.1 [Internet]. Document freely available on the internet at: http://www. r-project. org. Vienna, Austria; 2015. Available from: https://www.r-project.org/

44. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. shiny: Web Application Framework for R [Internet]. 2016. Available from: https://cran.r-project.org/package=shiny

45. Springer S, Wang Y, Dal Molin M, Masica DL, Jiao Y, Kinde I, et al. A Combination of Molecular Markers and Clinical Features Improve the Classification of Pancreatic Cysts. Gastroenterology. 2015;149(6):1501–10.

46. Ryu JK, Woo SM, Hwang J-H, Jeong JB, Yoon YB, Park IA, et al. Cyst fluid analysis for the differential diagnosis of pancreatic cysts. Diagn Cytopathol. 2004;31(2):100–5.

47. Van Der Waaij L a., Van Dullemen HM, Porte RJ. Cyst fluid analysis in the differential diagnosis of pancreatic cystic lesions: A pooled analysis. Gastrointest Endosc. 2005;62(3):383–9.

CURICULUM VITAE

The author was born in Karachi, Pakistan in 1990. He received high school and college education in the British GCE system at the Karachi Grammar School and went on to become a medical graduate from the Aga Khan University Medical College, Karachi, Pakistan, one of the leading medical schools of the country.

The author aspires to bridge the gap between medicine and the formal sciences (such as decision science and mathematics) by adapting techniques developed within the formal sciences into medicine. In doing so the author hopes to shed new light on the way doctors and medical researchers think about age old problems and uncover new roads to discovery.

The authors primary interests lie in surgical oncology and understanding important diagnostic and treatment decisions in cancer patients.